DOCUMENT RESUME

ED 336 417 TM 017 203

AUTHOR	Bradburn, Norman M., Ed.; Gilford, Dorothy M., Ed.
TITLE	A Framework and Principles for International Comparative Studies in Education.
INSTITUTION	National Academy of Sciences - National Research Council, Washington, DC. Commission on Behavioral and Social Sciences and Education.
SPONS AGENCY	Department of Defense, Washington, D.C.; National Center for Education Statistics (ED), Washington, DC.; National Science Foundation, Washington, D.C.
PUB DATE	90
NOTE	54p.
AVAILABLE FROM	Board on International Comparative Studies in Education, National Research Council, 2101 Constitution Ave., N.W., Washington, DC 20418.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Comparative Analysis; *Educational Research; Elementary Secondary Education; Guidelines; *International Studies; National Organizations; Research Design; Research Needs; *Research Proposals
IDENTIFIERS	*Board on International Comparative Studies Educ

ABSTRACT

A framework is presented for use by the Board on International Comparative Studies in Education (BICSE) in advising the National Center for Education Statistics and the National Science Foundation on United States participation in international comparative studies of education. The framework may also be helpful in identifying areas of research that are neglected. Also presented are principles recommended by the board for appraisal of proposals to conduct international education studies. The framework considers the value of United States participation in international studies and the kinds of studies that can be conducted. The measurement of educational achievement is reviewed, with attention to the many complications of international research. The long-term needs for United States participation and the timing and focus of proposed studies are considered. The recommended principles for appraisal of international studies apply to both explanatory and descriptive studies, whether of a qualitative or quantitative approach. Principles to be reviewed include considerations of relevance, design, analysis, and application. These principles have been adopted to guide the BICSE's appraisal of planned activities and for consideration by those interested in international comparative studies. Biographical summaries of 18 BICSE members are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED336417

A Framework and Principles for International Comparative Studies in Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

WILLIAM M. BURNS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NRC NAS NAE IOM National Research Council

BEST COPY AVAILABLE

A Framework and Principles for International Comparative Studies in Education

Norman M. Bradburn and Dorothy M. Gilford, Editors

Board on International Comparative Studies in Education

Commission on Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1990

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The project that is the basis for this report was supported with funds from the National Science Foundation, the National Center for Education Statistics of the U.S. Department of Education, and the U.S. Department of Defense.

Available from

Board on International Comparative Studies in Education
National Research Council
2101 Constitution Avenue, NW
Washington, DC 20418

Library of Congress Catalog Card Number 90-63855

Printed in the United States of America

BOARD ON INTERNATIONAL COMPARATIVE STUDIES IN EDUCATION

NORMAN M. BRADBURN (*Chair*), National Opinion Research Center and Department of Psychology, University of Chicago
JOSEPH F. ALIBRANDI, Whittaker Corporation, Los Angeles, California
GORDON M. AMBACH, Council of Chief State School Officers, Washington, D.C.
ROBERTO M. FERNANDEZ, Department of Sociology, Northwestern University
JAMES W. GUTHRIE, School of Education, University of California, Berkeley
EDWARD H. HAERTEL, School of Education, Stanford University
MORRIS H. HANSEN,* Westat, Inc., Rockville, Maryland
STEPHEN P. HEYNEMAN, Population and Human Resources Division, Europe, Middle East, and North Africa Region, The World Bank, Washington, D.C.
DANIEL G. HORVITZ, Research Triangle Institute, Research Triangle Park, North Carolina
LYLE V. JONES,* Department of Psychology, University of North Carolina
GAEA LEINHARDT, Department of Education, University of Pittsburgh
JOHN R. SCHWILLE, College of Education, Michigan State University
FLORALINE I. STEVENS, Los Angeles Unified School District, Los Angeles, California
JUDITH V. TORNEY-PURTA, Department of Human Development, University of Maryland
DAVID E. WILEY, School of Education, Northwestern University
RICHARD M. WOLF,* Teachers College, Columbia University

DOROTHY M. GILFORD, Director
E. LAURA LATHROP, Research Assistant
M. JANE PHILLIPS, Administrative Secretary

* Term ended June 30, 1990

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

PREFACE

The Board on International Comparative Studies in Education was established in 1988 by the National Research Council, through its Commission on Behavioral and Social Sciences and Education, to oversee educational research and statistical activities that are conducted in the United States in conjunction with other countries. The general purposes of the board are to develop periodically a comprehensive plan for U.S. participation in international studies; provide a forum for information and discussion; assist in planning the conduct and funding of studies; establish principles regarding the quality of study design, data collection and analysis procedures, and report preparation; assist in the dissemination of study findings; and promote the use of assessment findings to improve U.S. education. The board is currently funded by the National Science Foundation, the U.S. Department of Education, and the U.S. Department of Defense.

This document evolved from early activities of the board. As the board reviewed the plans for the Computers in Education Study being conducted under the aegis of the International Association for the Evaluation of Educational Achievement, (IEA), it became clear that guidelines should be developed for reviewing proposals and responding to agency requests for advice on whether to participate in specific studies. Thus, we began to develop principles for appraising proposals for international comparative education studies.

During its second year, our sponsoring agencies requested the board's advice on plans of the Educational Testing Service for a second International Assessment of Educational Progress, the second cycle of what might become a new series of international studies. Since the proposed series would compete for funding and for access to schools with studies planned by the IEA, which had been involved in international assessments for 30 years, the board recognized the need for a conceptual framework for a long-range plan for international studies. Therefore, we began development of a framework for advising government agencies on participation in international comparative studies. In addition to considering the question of U.S. par-

ticipation in international comparisons, the framework considers why the United States should participate in international studies and what kinds of studies it should support; discusses general issues in comparative education studies; and proposes a framework for establishing priorities for different types of studies. The framework may also be helpful in identifying areas of research that are neglected.

The board recognizes that it cannot unilaterally establish a framework for international studies, but it hopes to stimulate international discussion of such a framework. We offer this document as a basis for international discussion of the issues that must be considered in establishing priorities for different types of studies. The board plans to disseminate the document widely in the education research and policy communities, and welcomes comments on it. As a next step in building consensus, the board plans to convene a conference on the framework and principles in the spring of 1991 for discussion of the need for specific studies, including a desirable schedule for them. Following the conference we expect to prepare a prescriptive framework report that recommends a long-range plan for U.S. participation in specific international studies.

I want to extend my appreciation to the members of the two working groups who prepared the draft documents that were the basis of this report. The group of James Guthrie (*chair*), Edward Haertel, and Judith Torney-Purta developed the draft "Principles for Appraising International Comparative Education Proposals" and repeatedly revised it to reflect discussion at board meetings. The group of Stephen Heyneman (*chair*), Edward Haertel, Lyle Jones, Gaea Leinhardt, and Judith Torney-Purta drafted the "Framework for International Comparative Studies in Education." I also wish to thank all board members for the stimulating discussions that ultimately shaped this report.

The board sent the draft framework and principles documents for review to almost 100 comparative education researchers, including members of the general assembly of the IEA, several national research coordinators for the International Assessment of Educational Progress, numerous members of the European Consortium of Institutes for Educational Research and Development, and several education researchers in the United States and in international organizations concerned with education.

We received comment from 39 researchers in 19 different countries and 7 from researchers in international organizations. They are too numerous to mention by name, but we would like to thank all of those who gave so generously of their time in reviewing the drafts. They provided thoughtful and incisive comments, some of them based on the experience and knowledge acquired in life-time careers in comparative education. Their comments were of invaluable assistance to the board.

The board is grateful to Eugenia Grohman, Associate Director for Reports for the Commission on Behavioral and Social Sciences and Education, for her fine technical editorial work, which contributed greatly to the organization and readability of this document. We would also like to thank members of the Commission on Behavioral and Social Sciences and Education who reviewed the manuscript and offered cogent comments.

Finally, I want to extend my appreciation to Dorothy Gilford, director of the board, who is coeditor of the report and who provided the staff support that was indispensable to completion of the document and that made life tolerable for the chairman. As research assistant, Laura Lathrop was most helpful in classifying the hundreds of comments received from our colleagues who reviewed the preliminary drafts, and she handled all logistical arrangements for board meetings efficiently and effectively. Jane Phillips serves ably as administrative secretary for the board and cheerfully and competently coped with multiple rounds of revisions of this document.

Norman M. Bradburn, *Chair*
Board on International Comparative
Studies in Education

CONTENTS

A FRAMEWORK FOR INTERNATIONAL COMPARATIVE STUDIES IN EDUCATION 1

The Value of U.S. Participation in International Studies, 1
Kinds of Comparative Studies, 5
The Measurement of Educational Achievement, 6
Long-Term Needs for U.S. Participation in International Studies, 9
Timing and Focus of Proposed Studies, 11
Values to Different Constituencies, 12

RECOMMENDED PRINCIPLES FOR APPRAISING PROPOSALS FOR INTERNATIONAL COMPARATIVE STUDIES IN EDUCATION 14

Introduction, 14
Explanatory and Descriptive Studies, 15
Quantitative and Qualitative Studies, 16
Relation to Education, 18
Relation to Other Studies and Information Sources, 18
Relation to Policy, Practice, or Understanding in the United States, 19
Attention to Educational Influences and Cultural Context, 20
Conceptual Coherence of the Research, 21
Research Neutrality and Involvement, 21
International Scope, 21
Personnel, Institutional, and Financial Capacity, 22
Technical Validity, 23
Sampling and Access to Schools, 24
Content Sampling and Design of Achievement Items, 26
Coverage of Performance and Higher-Order Skills, 27
Instrument Construction, 27
Test Instruments, 27
Background Questionnaires, 28
Test Administration, 30

Plans for Analysis, Reporting, and Dissemination, 31
Analysis Plan, 32
Level of Detail in Reporting, 32
Standards and Criterion Levels, 33
Special Reports for Nontechnical Audiences, 33
Data Audit and Evaluation, 34
Public Use of Data, 35

BIOGRAPHICAL SKETCHES OF BOARD MEMBERS AND STAFF 36

A FRAMEWORK FOR INTERNATIONAL COMPARATIVE STUDIES IN EDUCATION

Investment in education is one of the principal means by which individuals and societies improve their well-being. The President and the governors have recognized the importance of education to the future of the country and have adopted a set of national goals for education. International comparative studies of education can assist school teachers and other professional educators, policy makers, the general public, and the research community in improving education in the United States and in measuring progress toward the realization of the national goals.

This document presents a framework for use by the Board on International Comparative Studies in Education in advising the National Center for Education Statistics and the National Science Foundation on U.S. participation in international comparative studies of education. The framework may also be helpful in identifying areas of research that are neglected.

This section of the paper covers four topics: the value of U.S. participation in international comparisons; kinds of comparative educational studies; measurement of educational achievement; and long-term criteria for U.S. participation.

The Value of U.S. Participation in International Studies

The most important reason for United States participation in international studies of education is to improve understanding of our own education system, that is, as an extension of and complement to studies within the United States. Since there are no absolute standards of educational achievement or performance, comparative studies are vital to policy makers in setting realistic standards and in monitoring the success of educational systems.

Through the use of standardized tests, school officials are able to compare the performance of their pupils with some external standard. Studies that compare academic performance

among schools within a single system provide information to school boards about the relative success of different schools; comparisons over time provide information about improvement or decline over the years. These comparisons, however, are limited by the nature of the reference groups or criteria used: that is, they are usually limited to school systems similar to those being evaluated. Even if schools are doing well when evaluated by local standards, how do boards know how well it is possible to do?

Comparisons with other localities are helpful. A natural comparison is with other similar local educational systems within the same state, or with those in other states or the nation as a whole. Such comparisons have been done at the national level for a number of years by means of the National Assessment of Educational Progress (NAEP). In 1990 data were also collected at the state level on a trial basis so that state-by-state comparisons can be made for participating states. Comparisons with other states or the nation as a whole have the advantage of comparing between educational systems that are broadly similar. But this advantage is also one of their limitations.

International comparisons expand the range of comparison beyond the parochial limits of the U.S. national experience. They provide information on the U.S. level of achievement in relation to the much broader range of the world's education systems. Recent international comparative studies, for example, have revealed that U.S. pupils could attain a much higher level of achievement in science and mathematics than they currently do. Collection of data at regular intervals from a large and diverse group of countries is thus important for descriptive or monitoring purposes.

Comparative studies can also be helpful in understanding the reasons for observed differences in performance. Studies that explore the relationship between school achievement and such factors as curricula, amount of time spent on school work, teacher training, classroom size, parental involvement, and a host of other possible explanatory variables profit from expanding the range of variation in such factors to the international level. While there is some variation in the characteristics of school systems in the United States, they are not radically different from each other. Schools in different parts of the United States

probably differ more in terms of the characteristics of their student bodies than they do in the ways they teach or in their curricula. Thus, the generalizability of the results from U.S. studies are quite limited. Careful international comparative studies can help identify the factors that promote educational achievement and those that do not make a difference. Such studies are difficult to do, however, because there is considerable uncontrolled variation in variables other than those of policy interest, which may make it difficult to reach sound conclusions.

International studies can also be important for issue-centered studies. Sometimes another country will exemplify a particular characteristic with special sharpness that makes it worth an intense comparative study. For example, the study of a country in Asia where student motivation in science and mathematics is extremely high, of a country in Europe where there is a high level of value homogeneity among schools, churches, and families, or of a country where employers participate actively in vocational education might provide insight into ways in which U.S. schools could be improved or into policies that are unlikely to work in the United States even though they may work well in another country.

In addition to their value as an extension of internal U.S. evaluation studies of education, international studies of education are important for subsidiary reasons. Instances are increasing in which having an American-only sample is inefficient for the purpose of developing improvements in the effective delivery of education. The issue here is not whether an observed pattern is typical, but rather whether something that exists in another country, but not in the United States, would be useful here. From one nation to another, education as an enterprise contains many similar exigencies and challenges: its methods of financial support; its role in determining what skills are provided at public and private expense; its mechanisms for treating the learning impaired and the socially underprivileged; its mechanism for rewarding excellence in teaching; and, ultimately, its decision as to what knowledge is most worth having.

These are not in any sense American challenges; they are universal. Both local and state U.S. education officials depend on a constant source of good ideas on which to base their management efforts. The range of ideas within a single district

usually is less than across many districts, less across districts than across states, and less across states than across nations. New ideas gained from international studies can be tried in the United States to see if they will improve the educational system. But it is critical to know more about input—the system variables and the goals for education—to understand why the performance of students in different countries differs.

International educational research also enhances the research enterprise itself. Many of the advances in the theory and practice of comparative educational research have come from innovations developed during international comparative studies in which the problems of comparison are most challenging. In many countries, working on an international research project helps to disseminate rapidly new models, new computer programs, and new statistical techniques. These in turn become debated locally, improved, and used again by those working on local problems.

Finally, international research records the diversity of educational practice. In any enterprise as diverse as education, there are practices and policies that deserve to be chronicled, not just on the grounds of their perceived utility, but on the grounds that they exist: for example, the number of languages taught in the classroom, the prevalence of pen and ink, the memorization of sacred texts, the use of Mark Twain as literature. It cannot easily be said that having information on these issues in different countries is likely to improve the practice of U.S. education, but it is worth knowing what exists in the world and, if the practices die out, what did exist but did not survive. Such knowledge may help educators avoid reinventing a faulty wheel.

In sum, international comparative research on education provides an important addition to research within the United States. It increases the range of experience necessary to improve the measurement of educational achievement; it enhances confidence in the generalizability of studies that explain the factors important in educational achievement; it increases the probability of the dissemination of new ideas to improve the design or management of schools and classrooms; and it increases the research capacity of the United States as well as that of other countries. Finally, it provides an opportunity to chronicle practices and policies worthy of note in their own right.

Kinds of Comparative Studies

Comparative studies of education can be arrayed on a continuum ranging from theoretically grounded studies intended to build or test complex models of educational systems to descriptive studies whose purpose is to monitor or document characteristics of educational systems, practices, or outcomes. More theoretically oriented studies tend to examine relationships among variables and look for causal explanations. For example, they might be designed to examine links between school achievement and such characteristics as curricula, teaching methods, family expectations, and funding levels. These kinds of studies are intended to identify influences on learning and how learning can be improved. They may focus on differences or amount of variation between schools or classes as well as on differences between students as the unit of analysis. These studies are expensive to conduct, but they are essential for policy makers and practitioners in their efforts to improve schools and the achievement of pupils.

Less theoretically oriented studies may only collect comparative data on test performance, curricula, school calendar, teacher salaries, or other indicators of the educational system. The goal of such studies is to provide useful, precise information on a few simple variables. The power of these studies lies with their rigorous sampling and, hence, their capability to make national estimates of the variables studied; the clarity of the findings; and the speed with which findings can be reported. The limitation is that they usually provide little or no data with which to interpret the reasons for observed differences. Many of these studies consist of educational information that can be periodically monitored: the level and variation of teacher salaries; the number and kind of available reading materials; and the level and variation in learning achievement in the more common subjects such as mathematics, science, and reading. Much of this information—enrollment, dropout rates, budgetary statistics, etc.—can be obtained from official sources and does not require special data collection, although there are often problems in the comparability of official statistics due to differing definitions of data elements. Other data, particularly those on academic achievement, must be gathered by special studies.

Descriptive studies measure trends over time, establish the range of variation that exists among countries, or chart the progress of educational reforms. They are of increasing interest to governmental policy makers as governments are more concerned with the relationship between investments in human capital and economic performance.

From time to time, special studies can focus on a problem, an issue, an exemplary program, or a contrast in educational policy or practice that can be illuminated by the study of schools in a small and selected group of countries. Sometimes countries will represent a "naturally occurring experiment"—for example, countries that have different teaching methods, countries that vary in the degree of involvement of parents, or countries that vary in their relation to particular employers or higher educational institutions. Issue-centered studies are likely to use a wider variety of methods than descriptive or explanatory studies and will sometimes take the form of case studies.

The Measurement of Educational Achievement

The term "educational achievement" is used to refer to skills, knowledge, and understanding that students acquire as a result of their participation in the educational programs of schools. Achievement is usually measured by some sort of test that may be—but often is not—related to the curriculum being taught in the schools the students attend. Studies of educational achievement may also be concerned with aspects of school systems that have some presumed relation to achievement, such as enrollments and dropout rates, as well as such characteristics of school systems as teacher qualifications, length of the school year, and amount of money spent per pupil. Informative studies of educational achievement often include attention to students' motivation to learn and to expend the effort necessary to perform well on tests.

There is no commonly agreed upon measurement scale for educational achievement analogous to the thermometer or the yardstick. In the absence of any common scale, the measurement of educational achievement relies on two strategies. The first is an explicitly comparative approach. A test is constructed whose content contains material on the knowledge and skills

that is thought to encompass the range of material taught in the grades of the students who are being evaluated. Then the tests are given, and norms are constructed on the basis of results obtained from those taking the tests. Results for individual students, classrooms, schools, or school systems are reported as numbers that are compared with the distribution of scores for the whole population taking the test. Thus, results would be reported for a student as "reading at the level of the average fourth grader," when the average for fourth graders is based on the empirical results of a large number of fourth graders who have taken the test.

An alternative approach is to decide on a level of knowledge that is expected to be achieved by the average student at some level of development, for example, the fourth grade. This level may be set by teachers, curriculum specialists, school boards, parents, or any group that has responsibility for evaluating educational outcomes. In this form of testing, results are reported as the proportion of items that a particular pupil answered correctly or the proportion of students in a particular classroom, grade, or school system that reached a designated criterion level: for example, "80 percent of the students in the fourth grade of a particular school system know the multiplication tables through 9."

The two approaches yield somewhat different information. The first, sometimes called "norm-referenced testing," shows how particular students or groups of students compare with a reference population, for example, fourth graders. The second approach, sometimes called "criterion-referenced testing," shows how much particular students or groups of students know in relation to a defined body of knowledge. These approaches are not mutually exclusive. Norms can be reported for criterion-referenced scales, and any well constructed test can be criterion-referenced by "anchoring" its scales if there are enough items and those items discriminate at various scale points. Neither strategy dominates educational evaluation today, although criterion-referenced testing, which is a newer approach, is becoming more popular. The Iowa Tests of Basic Skills are a familiar example of norm-referenced testing; the National Assessment of Educational Progress (NAEP) is an example of criterion-referenced testing.

Most achievement tests, whether norm referenced or criterion referenced, are multiple-choice paper-and-pencil tests. There is increasing interest in both the United States and other countries in performance tests, that is, tests that require pupils to do or write something that demonstrates their ability to solve a problem or perform an activity. For example, students might be given a ruler, a protractor, and a piece of paper and asked to draw an equilateral triangle. Performance testing is particularly valuable if it can become part of the ordinary classroom activity and is not seen as a separate and intrusive activity, as is often the case with achievement testing.

Progress has been made in developing performance tests that meet the measurement criteria necessary for valid comparisons, but further developmental work is necessary before they can be easily accommodated in international comparisons. Performance testing is also more expensive and makes more logistical demands on test administrators, which create further barriers to their widespread use. Nonetheless, the board believes that performance testing is a promising methodology that would have considerable value in both public and teacher acceptance of international comparative studies. We encourage its further development.

Another issue in measuring educational achievement concerns the way to select the students to be tested. Because the school years are also years of rapid physical and mental growth independent of any schooling, it is not clear whether students should be tested according to their age or to their years in school. Children start school at different ages; first graders may be 5, 6, or 7 years old, depending on their birthdays, the particular rules of the school system in which they enter school, and parental preference. Grade progression also occurs at different rates across countries. Some of the Nordic countries have policies against repetition. Thus, if one were interested in evaluating achievement at about the transition between "lower" and "middle" school, should one test fourth graders or 9-year olds? In comparing systems with different age rules for school entry, there may be quite large differences in the average age of students in the fourth grade. Again, there is no consensus on which strategy is most appropriate, and different testing and evaluation programs have different decision rules on this issue.

Population coverage is also an important issue in comparative studies. Countries differ in many ways: the legal age of students leaving school, the proportion of students dropping out before completing normal schooling, and the degree of channelling of students into different programs and types of schools. When comparing student performance among countries, it is important that the populations sampled are defined in similar ways and that there is comparable coverage of the populations.

The reliable and valid international comparison of educational achievement is not a simple matter. While the theory and methods of achievement measurement are well developed, their application in cross-national studies is neither straightforward nor easy. Such studies are among the most challenging that can be undertaken. They should not be undertaken without adequate resources for detailed planning, for data collection in each of the countries, and, especially, for comparative data analysis at the end of the study. It must be recognized that international comparisons are more expensive than simple comparative studies within one country. Given the importance attached to the results of international comparisons today, it is better to forgo a study altogether than to try to proceed with inadequate funding.

Long-Term Needs for U.S. Participation in International Studies

The board's concerns embrace the mix of international comparative studies in which the United States participates as well as the merits of particular studies. Generally speaking, comparative studies supported by the United States should address a range of content areas and grade levels and should encompass quantitative survey research studies as well as more intensive studies that use a range of qualitative research methods. Although most studies may be limited to paper-and-pencil measures of educational achievement, there is also a need for some studies that use performance tests.

International educational studies appear to be so important that the United States should plan to participate in the preparatory meetings, obtain the necessary commitments from local and state officials, and set aside sufficient resources to ensure

that the data gathering and analytic work will meet internationally recognized standards. The United States should also actively support methodological work designed to improve the reliability and validity of international comparisons. If international comparisons are to be technically valid and useful, issues of reliability and validity must be addressed outside the context of individual projects. There is a real need for more thoughtful, less constrained research on the methodology of international comparisons.

The United States should collect some information through a regular cycle of specialized studies. On a regular basis, the academic achievement of U.S. students in different subject matters should be compared with that of students elsewhere. While some data on variables that might have value in helping to understand obser ed differences in performance should be included, for reasons of cost and efficiency the major thrust of these studies should be simply to compare academic achievement of U.S. and other students. Such descriptive studies should be conducted frequently enough for policy makers to monitor changes in educational progress, but not so frequently that there is little likelihood of a detectable change.

Aspects of educational systems can also be monitored through a system of comparative education indicators. Much of the data for such indicators is already collected regularly by countries, although there may be a serious question about the comparability of indicators. The United States should participate in international programs, such as one currently being developed in the Organization for Economic Cooperation and Development (OECD), to provide education indicators on a systematic basis, but it must be recognized that the project faces both technical and conceptual difficulties.

Studies that explore influences on learning in some depth—by investigating such factors as details of school management, curricular diversification, classroom interaction patterns, community and parental influences, classroom material resources, or teacher quality—should be done from time to time as relevant theoretical models or significant new educational practices are developed. Every effort should be made to coordinate the administrative mechanisms for these two types of studies so that

duplication of effort is kept to a minimum and the opportunity costs for students and school officials is kept low.

The United States should also participate in issue-centered studies on particular problems about which other countries have a common interest. These special studies may not require national representative samples and need not occur on a regular basis. Examples of questions addressed by such studies are: Does classroom competition affect ethnic groups differently? What incentive programs are most successful in attracting the best science and mathematics teachers? How many classroom preparations a day are optimal for good teaching? Are outcomes better when employers finance industry-wide mechanisms for vocational training? There will always be a need for international information on issues of this kind. The timing and depth of analysis, however, should be determined by the level of need and the specific resources required for each problem separately.

Comparative studies take time away from classroom activity and may be seen as intrusive by school administrators and teachers. Good motivation to participate in the studies on the part of both students and teachers, however, is necessary if the studies are to be done well and the results are to be valid for each country. It is therefore important that every effort be made to develop studies that are useful to teachers and schools officials in improving the performance of schools, as well as useful to policy makers and researchers. Feedback to schools about the results of the studies and their implications for educational practice is also important.

Timing and Focus of Proposed Studies

Data collected over time in time-series or cohort designs can be of significantly greater value than single, cross-sectional studies, especially when data are collected at regular intervals. For that reason, high priority should be given to continued U.S. involvement in studies for which failure to participate would jeopardize valuable trend lines. Conversely, because it is difficult to make substantial alterations in the content or administration procedures used in data collected over time, the United States

should strive to ensure that studies intended to *initiate* a series represent the state of the art in design and instrumentation.

The scheduling of data collections embedded in a series is closely constrained. Flexibility in the timing of cross-national studies may also be limited by school calendars around the world and by the logistics of international cooperation. Nonetheless, the optimum timing of international studies should be considered in decisions to participate in them. Reasons for accelerating or delaying studies might include:

- Effects on the participation of nations and of sampled units within nations if too many cross-national studies are carried out simultaneously;
- The opportunity to evaluate specific, significant educational policies or investments in the United States or abroad;
- The expected impact on the diagnosis of major shortcomings of educational systems and on the development of remediating strategies and policies;
- Desirability of timing the release of findings to maximize impact;
- Documentation of educational systems or practices soon to be altered or eliminated;
- Likelihood that resources available for studies may be diverted to other purposes if there are undue delays, or conversely, that additional resources may become available at some future date.

Proposers of studies should also consider the potential overlap of any new study with other recent or ongoing studies. The utility of overlap for calibration of measures, comparison, and cross-validation must be weighed against the value of new distinctive data. The distinctiveness of a proposed study might be reflected in several key features: nations represented, academic content area, types of learning outcomes examined, age or grade levels involved, and research methods used.

Values to Different Constituencies

The primary factor in deciding on U.S. participation in comparative studies should be the information needs of the United States. In making the decision to participate, however, consid-

eration should also be given to a proposed study's value to other participants. As part of a global community, the United States cannot take an exclusively national view of any study's utility. The information needs of other nations, especially developing nations, may differ from our own, and the United States may sometimes be called on to join in studies that are of greater value to other countries than to itself.

Decision makers at different levels of the educational system have varying needs for information. Teachers and administrators at the school and district levels may seek information about specific instructional practices, while state and federal policy makers are more likely to be concerned with the effects of broad policies and programs. International organizations may wish to compare educational systems or evaluate development initiatives at the level of entire nations. At the national level especially, a study's importance may lie as much in drawing attention to an educational problem and catalyzing action as in providing new knowledge. Other things being equal, preference should be given to cross-national studies that address needs at more than one level. The aims and priorities of each study, however, should be clearly stated at the outset.

A proposed study's importance to constituencies other than those of the sponsoring agency should also be considered. In the light of the enormous economic importance of a sound educational system, leaders in business and industry may wish to consult comparative educational studies in their international planning. Textbook publishers, developers of educational software, and other educational vendors may use these studies to identify needs and markets for new products. Finally, if international research is to serve the end of scientific knowledge, it must be available to and used by the educational research community. Reporting and dissemination targeted to the needs of different audiences will enhance the value of an international study.

RECOMMENDED PRINCIPLES FOR APPRAISING PROPOSALS FOR INTERNATIONAL COMPARATIVE STUDIES IN EDUCATION

This section presents the principles recommended by the board for the appraisal of proposals to conduct international education studies. These criteria do not constitute a precise set of standards to be applied rigidly in assessing all proposals. Rather, they are the dimensions that the board believes should be considered in reviewing plans for international comparative education studies in which the United States is a prospective participant or contributor. Comparative studies that exclude the United States are obviously also important in the larger, global educational context of which the United States is a part, but the board is unlikely to review proposals for such studies. These principles have been adopted both to guide the board's own appraisal of planned activities and for consideration by all those who are involved or interested in international comparative studies.

Introduction

The board encourages the conduct of international comparative studies across a wide range of research strategies, formats, and procedures and a broad range of nations. In the past, many of the most widely publicized research efforts have been rooted in cross-national comparisons of student academic achievement. The dominant method has been item and student sampling, that is, collection of responses from each student for a sample of items from a pool and careful scientific sampling of schools or classes. Where appropriately conducted, this is a productive line of research and the board encourages similar efforts in the future. However, there are other research models, some highly quantitative, others relying on rigorous qualitative techniques, that also can enhance knowledge. The board also encourages international studies using qualitative techniques, especially when

they enrich or parallel previous or contemporaneous quantitative studies.

Explanatory and Descriptive Studies

Comparative education studies may be more or less directly grounded in educational models or theories. At one end of a continuum are theoretically based or explanatory studies intended to build or test complex models linking educational resources, practices, and outcomes. At the other end are descriptive studies, intended only to monitor or document critical facets of educational systems, practices, or outcomes.

More theoretically grounded studies often probe the relationships among variables in an effort to seek evidence for causality. For example, they might be designed to study the educational effects of cultural and other large contextual differences among countries or to determine the degree to which teacher characteristics, family expectations, textbooks, or funding levels are correlated with and might explain educational achievement. They might relate the education levels of different nations' populations to their financial support for schooling or to voter participation. They may also be designed to compare pedagogical approaches and their effects on students' learning by including longitudinal item-level data. Less theoretically oriented studies might include collection and compilation of data on student achievement, teacher salaries, curricula, or enrollments. They might map the range of variation, determine trends over time, or chart the progress of reforms. These studies are of increasing interest to policy makers as nations intensify their investments in human capital because they provide information that can assist in shaping and selecting from broad policy options. We caution, however, that the comparability of the results of such studies depends on the degree of similarity between the country contexts, and therefore the results must be placed in a clearly identified context.

In discussing the board's principles for appraising comparative education studies, we refer to less theoretically oriented studies as descriptive, and those that are explicitly grounded in particular theories as explanatory. We use the term explanatory

because explanation is the goal. However, it needs to be emphasized that correlations are not necessarily—and often are not—indicators of cause and effect. In addition, there is no sharp division between these two categories of studies, and any particular study is likely to partake of both purposes in some degree.

Quantitative and Qualitative Studies

Comparative studies also vary in their reliance on objective measurement, quantification, and narrative description and on use of statistical methods or systematic observation. There is no sharp division between these latter two research approaches, but we refer to the first approach as quantitative and the second as qualitative. Some studies use both quantitative and qualitative methods; in fact, qualitative strategies can be embedded in quantitative studies to illuminate relationships.

Quantitative studies most often rely on scientific samples from carefully framed populations that are usually defined at the level of individual students, although primary and intermediate sampling units may be at some other level of aggregation. Numerically quantifiable data are collected, usually with tests or questionnaires, and these sample data are used to support statistical inferences to the population. Quantitative methods can also be used to study resources, activities, and outcomes at the classroom or school level.

Qualitative studies are more likely to use samples defined at the level of classrooms, schools, or school systems, rather than individual students. The number of units sampled is typically much smaller than for quantitative studies, but they are investigated much more intensively. The sites investigated are usually chosen systematically to represent a range of demographic characteristics, organizational arrangements, or other features relevant to the questions to be addressed. Observations and interviews will be conducted over a period of time, sometimes by an investigator who participates in the ongoing activities of the school or other setting studied. Case studies can be used initially to document relationships that, once understood, can then be translated to survey formats; and survey results, in turn, can stimulate in-depth case studies. A special type of qualitative study is docu-

mentation relating to the history of education systems. Historical studies are very important for understanding the conditions that account for particular structures of schooling and achievement levels and can aid in developing realistic policy alternatives.

The fundamental principles of sound research apply equally to qualitative and quantitative studies, but there are different canons of systematic inquiry for each which entail different warrants for generalization. Thus, proposals for qualitative or historical studies and those for quantitative studies must be evaluated by somewhat different criteria.

In characterizing studies, other distinctions can also be made. Many studies are cross-sectional, obtaining data for only one point in time. Others are longitudinal, obtaining information on the same sample at various points in time, for example, at the beginning and end of the school year. Other contrasting approaches are large-scale, randomized surveys of entire nations versus smaller, localized, but intensive observational studies.

The board believes there is value in all these different varieties of inquiry and does not hold any particular research strategy, descriptive or explanatory, quantitative or qualitative, longitudinal or cross-sectional, to be uniformly superior. Rather, the overriding concerns are that the methods used be appropriate to the questions posed and that, regardless of topic or technique, a proposed study adhere to appropriate canons of systematic inquiry, consistent with the principles, enunciated below.

These principles are to be regarded as a set of basic standards to which proposed studies should aspire. Rather than suggesting *what* ought to be studied or *which* proposed studies would be of greatest significance, these criteria only suggest *how* a study ought to be conducted or what questions most proposals should address. In practice, of course, discussions about "how" will be shaped by views about what ought to be studied and the significance of the issues.

Finally, it will be clear that not all of these principles are relevant to all studies. Many pertain only to particular purposes or methods of inquiry. Moreover, many of the principles describe ideals that may sometimes be difficult or impossible to attain. Because of practical constraints imposed by time, resources, knowledge, and the sometimes competing values and interests of study participants, the design of every study must embody

compromises. Depth may be traded for breadth, sample sizes may be smaller and instruments shorter than ideal, and so on. There is probably no perfect proposal or perfect study. Consequently, researchers are encouraged to consider which principles are most relevant to their own investigations and to view these principles as ideals to strive for as they inevitably balance competing demands and practical constraints. Certainly all principles should be carefully considered in the design of any study.

Relation to Education

The board interprets "education" broadly. In addition to formal instruction delivered through various institutions to individuals of all ages (including adults), the term is intended to include activities, whether formal or informal, that directly relate to education and educational agencies and institutions. Areas within the purview of the board include studies or surveys of student performance or other educational outcomes; educational requirements; planning processes; curricula; instructional materials, resources, and practices; structural arrangements; professional preparation; parents', pupils', and professional educators' attitudes; enrollment and dropout rates; and those that analyze education as part of the political agenda or the economy. Even this list is only illustrative; it is by no means exhaustive. By way of contrast, proposed international comparative studies or surveys of the effects of nutrition, housing, or health effects on schooling, however significant and useful, would probably not be construed primarily as studies of educational activities, agencies, or institutions.

Relation to Other Studies and Information Sources

The value of achievement scores and other educational data or findings may be enhanced when they can be compared directly with information collected in the past or from other populations. Thus, the board supports the idea of studies that provide for linkages to earlier comparative studies or surveys in the same subject area, even though it recognizes that most international studies to date have not been so designed. Because of the

technical difficulties associated with monitoring trends over time, an appropriate statistical model should be a key design feature of such studies. When appropriate and feasible, the value of a proposed study may also be enhanced by the use of test items and data collection strategies that permit linkage to planned or ongoing national or regional data collections. Such linkages might be accomplished by providing for a core data collection with options for national augmentation. However, any such scheme should strive to ensure that augmentation does not compromise the validity of the international comparisons.

Relation to Policy, Practice, or Understanding in the United States

A proposal for an international comparative education study or survey should be appraised first and foremost on its likelihood of informing educational policies, practices, or the scholarly understanding of professional educators and researchers. Organizations and individuals planning such studies should not assume that the utility of what they propose is automatically evident. Thus, a proposal should include a list of the questions the proposers expect to answer, and it should include a description of its significance for informing policy makers, improving practice, or systematically adding to knowledge. In documenting how a critical issue will be addressed, the proposal should show inputs that can be manipulated by policy makers. It should show sensitivity to questions important to policy makers, administrators, teachers, researchers, and other stakeholders, and it should specify the means by which the analysis and study conclusions will be disseminated to relevant audiences in participating nations.

The board notes that studies narrowly limited to comparing highly aggregated mean levels of educational achievement for participating nations, assessed at a single point in time, are likely to be somewhat more difficult to justify in terms of their relevance to policy, practice, or understanding than are studies with the potential to illuminate the role of educational factors (e.g., organization of the curriculum or teacher training) in promoting achievement. They do, however, provide important contextual information for policy makers, particularly on

macrolevel and alterable variables. Clearly, the board has specific and particular concerns with the utility of cross-national studies to audiences within its own nation and therefore encourages proposals for studies of potential value to educational practice, policy, and research in the United States.

Every country's curriculum is rooted in its culture. Sometimes, in the interests of expanding a study to make it a wide-ranging cross-national comparison of achievement, data relevant to national understanding and national policy may be compromised. More detailed and purposeful studies of a small number of comparable countries may be more useful in these cases than large-scale cross-national studies.

Attention to Educational Influences and Cultural Context

The cultural context for learning may contribute to differences in expectations that affect not only what is taught but when it is taught. The fundamental problem of cross-cultural comparisons is the need for a strong theory explaining the contextual differences among the nations.

A proposed international study should display sensitivity to the cultural contexts (e.g., language spoken, religion, laws, implements used, values held) for the education dimensions to be assessed. The study plan should be reviewed by an individual in each participating country who understands how educational influences and cultural context shape and are shaped by policy. Also of concern are demographic and economic trends disaggregated by occupational divisions or rural-urban residence, for example, to permit examining the educational attainment of various subpopulations across nations. Among other concerns, the utility and interpretation of the study should be considered in the light of participating nations' resources, curricula, graduation requirements, and school-going populations. Even descriptive surveys, intended to chronicle the conditions of two or more nations on one or a few dimensions (e.g., teacher salaries or 12-year-olds' mathematics knowledge) should strive to provide information regarding the context—country wealth, value placed on technology, and so on—in which such conditions are embedded in each of the nations included in the sample. Although much of this information is available, organizing it into a com-

mon framework with interpretive usefulness can be very difficult.

Conceptual Coherence of the Research

Another underlying principle in considering proposals, particularly those for explanatory studies, is the degree to which the prospective study represents a conceptually cohesive research endeavor. This means that a proposal that is technically sound but that largely ignores past studies or is disconnected from existing bodies of knowledge in the study area, or in which intellectual elements of the research are fragmented or contradictory, may be inadequate. Descriptive studies should likewise demonstrate awareness of any recent closely related studies.

Research Neutrality and Involvement

An international comparative education study must avoid political, national, religious, racial, gender, or ideological bias. It is particularly important to make certain that, if western paradigms are used, they are relevant to other geographic areas. Therefore, it is essential that all nations to be included in a study participate in the study design, and mechanisms for facilitating such participation should be described in the proposal. Although it is important to safeguard against biases, actual differences (political, ideological, gender, and even religious) present challenges in comparative research that must be recognized. Such differences are often meaningful sources of cultural variation.

International Scope

Prospective studies submitted should have a clear cross-national scope, and the United States, either in toto or in appropriate states and regions, should be included among the nations proposed to be studied. The United States and at least one other nation should be involved, unless a study has already been done in the United States and the same study is being repeated in other countries to obtain relevant comparisons. In general, there should be no upper limit on the number of international comparisons to be undertaken, although for reasons of resources

and manageability it may be important to limit the number of countries participating in any given study. Involvement of developing countries in international studies contributes to the development of local research capacity and also broadens the sample of participating countries. Third-world participation improves North-South dialogue as well as East-West linkages. Education research studies are good vehicles for building trust and cooperation. The important consideration is that the proposed study be clearly cross-national in its scope and intent. Conditions under which countries (or national data) will be excluded from a given study—which are usually associated with data quality or failure to meet deadlines—should be made explicit.

Personnel, Institutional, and Financial Capacity

Organizations and individuals proposing a comparative international study should have qualifications and credentials appropriate for the proposed undertaking. The institution proposing the study or serving as the international center should demonstrate that it has a good research record, preferably in international research. The institution must show that it possesses among its staff the necessary organizational, language, psychometric, statistical, probability sampling, data management, and specific subject-matter skills, as well as staff who have a thorough knowledge of the principal ideas behind the educational systems that are included in the study and experience working with researchers in different countries and cultures. The individuals who coordinate the study within individual countries are also key for success of the study. They should have a very thorough knowledge of their own educational systems and of the subject areas under study, and they should have some experience with survey research. To participate effectively in the international planning meetings they need to speak the international common language which currently is English. Cross-national study organizers need to ensure that participating nations have available sufficient expertise to enable them to fulfill their obligations.

In addition to ensuring that the researchers involved possess the appropriate background and training, evidence should be

provided that financial resources being sought for the proposed study (or, occasionally, already available) are sufficient to conduct the study in a technically valid manner. The matter of sufficient resources is particularly significant. Past experience suggests that proposed studies are frequently well conceived, but that they later develop operational flaws due to debilitating compromises necessitated by inadequate resources. International studies cost more than national studies, but without realistic funding neither the quality of the work nor adherence to time schedules can be guaranteed. The board encourages organizations that are planning international studies and researchers who undertake responsibility for a country's participation in a study to avoid such situations by ensuring from the outset, to the extent reasonable, that adequate resources exist or will be obtained. Prior to undertaking a study, the organization responsible for the international aspects of the study should have firm funding commitments for international planning (both theoretical and operative); coordination; instrument development; training; data cleaning; analysis; and data documentation, preservation, and dissemination.

The study plan should demonstrate that the steps of the study are well integrated and mapped out in advance. Provision should be made for an initial task force to secure pertinent expert advice, and sufficient time should be provided to secure funding from multiple sources. Schedules and budgets should be realistic and should cover data analysis, reporting, and dissemination as well as study design and data collection. Finally, it is important to ascertain whether a proposed study is overly ambitious. Would participating countries have the personnel and financial capacity and endurance to complete a study with large numbers of instruments and questions, which would take up to 7 years, or would a more modest study be more productive in the long run?

Technical Validity

A complex education study may serve a variety of descriptive or explanatory needs, but its primary justification is likely to rest on the few central questions or issues it is designed to address. For any proposed international study, these key ques-

tions or issues should be explicit. In an explanatory study, the relationship of the issues to existing knowledge should be clear, and the study should be technically capable of addressing those issues. The proposed methodology, design, and statistics should fit the underlying model. The more specific guidelines that follow are subordinate to this general principle. Their importance to any particular study will depend on the major purposes the study is intended to serve. They are directed primarily to cross-national student achievement studies, which have been the focus of most of the board's early activity. The board's scope of activity is expanding and later revisions of the principles will include specific guidelines for other kinds of comparative studies of education, for example, studies that attempt to explain how differences in attainment are produced or those that focus on more culture-bound factors.

Sampling and Access to Schools

Nearly all quantitative studies, both descriptive and explanatory, as well as some qualitative studies, necessitate drawing a sample from the full population of all respondents, that is, all teachers, all administrators, all students at an age or grade level, or all policy makers. Valid estimation of population parameters from sample data depends critically on rigorous adherence to an explicit sample design. Whenever statistical inference from a sample to a population is intended, proposals for international comparative studies should describe in appropriate detail their plans for framing and selecting samples in participating countries as well as for exclusion of particular subgroups (e.g., persons who are developmentally disabled or who do not speak the language of the test). Subgroups should not be excluded solely for convenience in administering a test: for example, students not in the modal grade for the target population should not be excluded. Whenever a subgroup is excluded, information should be provided on the portion of the target population excluded and the extent and direction of bias introduced by the exclusion. Potential differences in student demographics among countries must also be considered. The population of students in countries in which the rate of participation in education is low may

be very different from the population sampled in a country where the participation rate is high.

Each sample should be designed so as to support reasonably accurate inferences about an age or grade cohort, and the proportion of each cohort covered should be carefully estimated and reported. The sample should be designed to ensure it captures the range of individual, school, or classroom variation that exists in the nation sampled. Explicit delineation of the populations and subpopulations to be sampled is critical. Within-country samples may be defined according to geographic regions, language groups, school systems or sectors (e.g., public versus private), or other relevant stratification variables.

The board recognizes the difficulty of defining comparable samples across different nations' school systems and curricula. Nonetheless, corresponding national samples should be defined in such a way that valid and informative cross-national comparisons are possible. To facilitate the sample selection, an international sampling manual is essential. In view of the complexities in this area, the board encourages the appointment of an experienced and expert sampling consultant to scrutinize sampling plans in all participating countries. Individual country samples should be approved by the international sampling consultant before testing takes place.

Well in advance of the date for test administration, arrangements should be made with the appropriate organization or individuals (ministry, state, district, school, teachers) to ensure high participation rates in the study. While the principle of strict adherence to an explicit sample design is sound, the *achieved* sample in actual international studies is usually different from the *designed* sample, especially so in countries in which response rates are low. The sampling manual should include a maximum acceptable nonresponse rate for inclusion of a country's data in the international analyses.

Subnational or regional units smaller than a nation should be allowed to participate in international studies if they have separate autonomous school systems. However, study results for such units should be reported in separate tables from the data for whole nations.

Even though the sample designs for large-scale studies satisfy the criteria described above, typically they cannot afford

the close direct observations that qualitative educational researchers want. Smaller in-depth studies of relatively small, localized samples in a small number of sites can also play an important role in comparative education research and policy development.

Content Sampling and Design of Achievement Items

Achievement items in an international comparative study may be used to support inferences about broad curriculum areas. Thus, it is critical that they be chosen according to an explicit and justified plan. The curricula of all participating nations should be considered in formulating such a plan, and content specifications should be developed through a consensual process involving representatives from all of the nations involved. Ample time should be allowed for meetings on content sampling and design of achievement items. At these meetings, information should be available on the purpose of each item, to assist the country representatives in selecting those that will evaluate the most important knowledge and skills. In general, coverage should be broadly inclusive. It will probably be desirable to assess a core of learning objectives common to most participating nations, but if there is general agreement on the importance of relevant, measurable learning outcomes that do not appear in participating nations' curricula, they may be included. It may also be desirable to include objectives in other domains, for example, student attitudes, values, and creativity. Matrix sampling (i.e., dividing the items to be included into subsamples to be administered to different students) might be considered as a means to increase the number and diversity of test questions included without unduly burdening individual survey respondents. The validity of test items should be reviewed by teams of experts that include cognitive scientists, educational psychologists, and curriculum or methods specialists in the relevant disciplines. The board recognizes the complexity of sampling curriculum content and the intractable problems of interpretation when comparing student outcomes for countries with very different learning objectives.

Coverage of Performance and Higher-Order Skills

When assessing student performance, objective questions can offer considerable assessment efficiencies relative to free-response items (such as open-ended questions), and multiple choice paper-and-pencil items can be designed to measure some higher-order skills. Nonetheless, consideration should also be given to the inclusion of test items and other data collection formats offering opportunities for students to display their performance abilities. Increased emphasis should be placed on writing, speaking, and interacting in both practical and school tasks. For example, reading, writing, and problem solving might be assessed in the context of particular subject areas. When feasible, complex, conceptual knowledge, process skills, and higher-order thinking should be assessed, as well as important factual knowledge, basic skills, and other outcomes usually achieved earlier and considered prerequisite for higher-level learning. Of course, there are economic considerations that must be taken into account in any study that uses "hands-on" assessment activities, but in most cases time and resources should be reserved to make some open-ended tasks possible.

Instrument Construction

Test Instruments

There may be sound reasons to use existing test instruments in international comparative studies, including continuity with earlier studies and linkage to other ongoing studies, as well as economy and efficiency. When new instruments are developed, however, they should adhere to high standards. Test content should represent a reasoned balance among the curricula and the information needs of all nations to be included in a study. The test development process should allow for participation by representatives of the various nations involved and should be informed by expertise in the curriculum area assessed, in the cognate academic discipline, and in educational measurement. Care should be taken to avoid redundancy among the questions. If new measures are proposed, there should be evidence

that the measure works in at least one country before it is included in an international study.

Whenever corresponding tests in more than one language must be prepared, the test should include some items originating in each of the languages represented. Consideration should be given to the development of parallel text materials that are constructed simultaneously within the cultural context of the different nations, rather than simply translated. If this is not feasible economically, and translation is used, all exercises should be back-translated to enhance accuracy and comparability. In addition, qualified bilingual experts should scrutinize pairs of tests, item by item, for unintended differences in emphasis or levels of abstraction. Care must be taken to ensure the equivalence of meaning of an item in the different languages.

New or substantially revised tests should be pilot-tested to ensure the quality of individual items and instructions to examinees, as well as the appropriateness of time limits for the questionnaire. Following the pilot test, a check should be made for item bias, including cultural bias or translation bias, by examining the relative difficulty of an item to other items in a subtest or domain. A check should also be made of the appropriateness of any statistical model used for scaling to ensure that it can cover the total range of scaled scores from all countries before the tests are used in any main testing.

A standardized research design across countries is essential, although national or international options can be added. Other modifications of the standardized design should not be permitted, since they can have serious consequences for validity or comparability.

Background Questionnaires

Educational achievement data cannot be appropriately interpreted in the absence of information about responding students, their backgrounds, their motivations, and their educational experiences. For cross-national studies of achievement test scores, it is especially critical that such information be collected. Background questions should be selected judiciously, and particular attention should be given to matters such as variables (a) relevant to the interpretation of achievement patterns, (b) plausibly related

to school achievement (including locally available educational resources), or (c) reflecting additional schooling outcomes valued in their own right.

Explanatory studies that rely on quantitative data should generally not rely exclusively on students' own reports of such factors. Such studies should also include instruments directed to teachers, administrators, and parents. For example, teachers or curriculum coordinators might be asked about the availability and use of particular instructional materials, local curriculum, or specific instructional practices.

A structural model that postulates cause-effect relationships to account for variation in student achievement should be used in selecting background questions. The model can also guide the analyses directed to identifying the sources of individual and group differences in achievement and the relative impact of these sources. Background variables about students seek to explore the relationship between students' background and home environments and achievement and attitudes. For example, information might be requested about the students (age, gender, race or ethnicity), indicators of family environment, parental encouragement, and attitudes toward school assignments in the subject matter being assessed. Information sought from teachers might include information about their teaching experience, availability and use of particular instructional materials, local curriculum, and classroom environment. School administrators might be asked for data on school factors believed to influence student achievement, such as instructional time, student enrollment and attendance, and programs in the subject area.

Background information collected from students, teachers, and school administrators can be supplemented by data from other sources that provide economic and social indicators for the various nations participating in the study. Economic and social indicators can be related to student achievement in various sectors of the population (e.g., rural or urban) and can also be used to explore the relationship of student achievement to economic development, resource development, industrialization, political stability, and the like across nations.

Representatives of all the countries participating in a study should be involved in developing background questionnaires as they are for the test instruments. Similarly, care should be

given to translation, back-translation, and scrutiny of background questions to ensure the equivalence of meaning of a question in different languages. The background questionnaires should be pilot-tested.

Because background data become more valuable if they can be compared over time and across populations, the same wording should be retained from study to study. Although it is difficult, effort should be made to ensure that background variables are defined similarly in the languages of all participating nations. Similar effort is required to ensure the comparability of social and economic indicators for all participating nations. All variant definitions should be documented.

Test Administration

Whenever achievement results are to be compared from one test administration to another, it is imperative that administrative procedures be controlled to be as nearly identical as possible. Maintenance of standard test administration procedures over time and from one nation to another is of paramount importance. Standardized procedures for instructing students and establishing conditions for testing should be developed, based on a pilot test of the instructions in each participant country. Time should be allotted at an international meeting of study coordinators to listen to their complaints and suggestions following the pilot test and to agree to standard administrative procedures. Testing materials should be clearly understandable. The testing environment should be comparable from one setting to another—both within and across nations—and should be free from distractions.

Each study design should address plans to control and standardize conditions of test administration. Ideally, to ensure adequate quality control, suitably trained people from outside the schools should be in charge of the test administration. In addition, people from different countries should supervise the implementation of the procedures to be followed (previously agreed on by the countries involved) by being present on site when the field work is conducted. Such quality control procedures would assure more uniform test administration, particularly in countries with little experience in assessment. Each design

also should address the level of student motivation to try to minimize any plausible systematic differences from one nation (and from one test administration within a nation) to another in incentive to perform well in response to test questions. Each country report should carry a description of test administration conditions.

Plans for Analysis, Reporting, and Dissemination

Plans for analysis, reporting, and dissemination of international comparative study findings should be described at the time the study is proposed and should indicate how the critical questions to be informed by that study will be addressed. These plans should provide for balanced reporting of cross-national comparisons and may also involve separate analysis and reporting of data from each participating nation or subsets of them. The board discourages exclusive, or even heavy, reliance on overall national rankings. Very often differences in educational systems render such comparisons invalid; a more productive approach is to find out the reasons for observed differences in pupil achievement. Prior to the release of any cross-national report, opportunities should be provided to all nations for review of the analysis and interpretations.

Without dwelling on them too much, reports should give prominent place to a discussion of the known and surmised limitations. Reporting should be sensitive to contextual factors that might affect test validity, for example, the relative familiarity of children in different countries with testing in general or with the particular item formats used in a comparative study. The possibility might also be considered that children who are exposed to a great deal of testing may expend less effort on "low stakes" tests they know do not matter for their own educational futures.

Reporting should also be sensitive to technical limitations on a study's interpretability. Limitations might include caveats about the comparability of national samples, the limited number of test items or range of content on which comparisons are based, differences in administration conditions from place to place, the match of tests to different curricula, the difficulty of translating exercises from one language to another, the limited precision of sample statistics, or other qualifications on study findings.

Analysis Plan

For various reasons, data analysis plans may change or evolve from the time a study is designed to the time it is completed and reported. Unforeseen difficulties in data collection or limitations of data quality may preclude some planned analyses. New questions or insights that occur in the course of data collection and analysis may open productive new lines of inquiry. Data already collected may be pressed into service to address emergent policy issues. Even when such evolution is anticipated, however, every proposal for an international educational study should include an analysis plan. The correspondence between the analyses proposed and the questions they are intended to answer—if not obvious—should be made explicit. In both explanatory and descriptive studies, it should be clear how theoretically central variables are to be measured and how relationships among critical variables are to be assessed. In qualitative studies, methods of examining and relating alternative data sources should be indicated, and anticipated procedures for developing conceptual or explanatory frameworks should be described.

Level of Detail in Reporting

In any complex study, there is a tension between the level of detail and the precision of the reported results. At one extreme, an average score over a large number of test items for an entire nation may be estimated quite precisely, but it conveys little information. At the other extreme, reports of numerous quantiles of the score distributions for narrow student subpopulations on individual items may be so poorly estimated that they also convey little information. However this tension is resolved, it is crucial that standard errors be calculated and reported with all reported statistics. Calculation of standard errors is technically complex, and the board encourages the use of a recognized expert consultant in this and other analysis stages, as it does for sampling.

The first issue to be resolved with respect to the appropriate level of detail in reporting is the number and size of subpopulations to be distinguished. Performance may be reported for major subgroups of student cohorts, defined by geographic region,

language background, gender, ace and ethnicity, or other variables,
if such reporting advances e purposes of the study. When
achievement is reported, the utility of multiple scores should
be considered. In many cas s, interpretive emphasis is prop-
erly given to major content nd process categories rather than
to total scores. Finally, with n the limits on precision imposed
by the design and size of a study, distributional summaries
should be given and not jus means and standard deviations.
Reporting of quantiles (e.g., eciles, or quartiles) is one method
that is readily explained and understood, and graphics such as
box plots are easily underst od and of potential value. Con-
sideration may also be giver to reporting at multiple levels of
aggregation if that is appro riate to the design and intent of
the study. In addition to esenting the student-level score
distribution, for example, di tributions of classroom or school
means might also be reporte .

Standards and Criterion Le els

Studies concerned with st dent achievement data can be en-
hanced considerably by reporti g outcomes in terms of performance
standards, for example, the ercentage of students who know
everyday science facts or w o use scientific procedures and
analyze scientific data. Th ; can be difficult to accomplish,
however, and there is a risk that arbitrarily established stan-
dards will lead to serious isinterpretations of achievement
levels. If results are reported relative to specified performance
levels (e.g., functional literac), the basis for establishing these
levels must be explicit, defe sible, and responsive to the needs
and contexts of all the natio s involved. This might imply the
use of different criterion leve s for cross-national reporting than
for national reporting. Alte natively, a graduated series of
proficiency levels might be efined, labeled with appropriate
descriptors, and illustrated w ith representative test items.

Special Reports for Nontech ical Audiences

Special reports should be prepared for nontechnical audi-
ences, including the press, po ticians, and policy makers. These

reports, which are designed to serve political purposes, differ from the more detailed reports intended for research and educational purposes. They should be designed so that the information is easily assimilated. Useful analytic tools for such reports include simple graphs, percentiles, and a graduated series of proficiency levels with illustrative test items.

Preparation of this type of report plays a role in institutional capacity building by forging links between the research and policy making communities. It also augments the dissemination of the latest information and techniques and will enhance long-term funding prospects. Study proposals should provide for mechanisms to disseminate results widely among public and private organizations. Such dissemination stimulates debate, which makes it more likely that study findings will be put into practice.

Data Audit and Evaluation

Experience has shown that national researchers make many changes in background questionnaires from the intent of the international questions. This leads to nonconformity of data to the international code book, which requires extensive work by the international coordinators, to clean the data. In some cases it is desirable to produce a data-entry program and a data-cleaning program for the use of national research coordinators.

The technical features of any international comparative study should be clearly documented. It is desirable that at least a summary of the methods involved be included in the principal reports, along with estimates of sampling precision. More detailed documentation, which might be published in a separate volume from the main report of the study, should address such matters as maintenance of the security of test materials before the actual testing; sampling adequacy (participation rate, attrition, absentee follow-up); comparability of administration conditions; procedures for audit of data collection; data checking, cleaning, and scoring; procedures for review of study reports prior to publication; and other procedural matters that may condition the confidence placed in study findings.

Public Use of Data

Countries participating in studies should be authorized to release their own findings as soon as the national data file is cleaned, merged into the international file, and ready for analysis. Provisions should be made to ensure that, when appropriate and within a reasonable period after analysis and reporting by project sponsors, data are placed in the public domain in a form accessible for secondary analysis. Special attention should be paid to making the data accessible to researchers in third-world countries. Clear and complete data documentation is crucial. When feasible, consideration should be given to using existing archives.

The importance of making international data easily accessible for secondary analysis should not be underestimated. More extensive use of the data at the national policy level can help in understanding the weaknesses and strengths of the U.S. educational system as well as those of other countries.

BIOGRAPHICAL SKETCHES OF BOARD MEMBERS AND STAFF

JOSEPH F. ALIBRANDI is chair of the board of directors and chief executive of Whittaker Corporation and a director of Security Pacific Corporation and Security Pacific National Bank, Santa Fe Pacific Corporation, Catellus Development Corporation, and Jacobs Engineering Group. Prior to joining Whittaker, he was associated with the Raytheon Company for 18 years, lastly as senior vice president. He is the former chair of the California Business Roundtable Task Force on Education. He also served as chair of the Twelfth District Federal Reserve Bank and chair of the International Policy Committee of the U.S. Chamber of Commerce. He is a member of the advisory board of Policy Analysis for California Education, cochair of "Kids First," and director of the Institute for Contemporary Studies. He received a B.S. degree in mechanical engineering from the Massachusetts Institute of Technology and a D.S. degree in business administration from Bryant College.

GORDON M. AMBACH is executive director of the Council of Chief State School Officers, which represents state-level interests in education and provides advocacy for their state and national education policy positions. Previously he served for 10 years as the New York State Commissioner of Education and President of the University of the State of New York. He also held appointments with the U.S. Office of Education in legislation and program planning. His advisory roles include membership on commissions and panels on learning technology, job training, the arts, and education statistics and assessment. He is currently the U.S. representative to the General Assembly of the International Association for the Evaluation of Educational Achievement. His professional interests have centered on policy making and legislative development. He has a B.A. degree from Yale University and an M.A. degree in teaching from Harvard University.

NORMAN M. BRADBURN is director of the National Opinion Research Council, the Tiffany and Margaret Blake distinguished service professor in the department of psychology, and a member of the faculties of the graduate school of business and the Harris School of Public Policy Studies, all at the University of Chicago. Previously, he served as provost of the University of Chicago. A survey methodologist, Bradburn chairs the U.S. Board on International Comparative Studies in Education and is a member of the Committee on National Statistics at the National Research Council. He is president-elect of the American Association of Public Opinion Research and is a fellow of the American Statistical Association, the International Statistical Association, and the American Association for the Advancement of Science. Bradburn received B.A. degrees from the University of Chicago and Magdalen College, Oxford University, and M.A. and Ph.D. degrees in psychology from Harvard University.

ROBERTO M. FERNANDEZ is associate professor of sociology and a research faculty member of the Center for Urban Affairs and Policy Research at Northwestern University. Previously he served as assistant professor of sociology at the University of Arizona. His research interests include the areas of formal organizations, race and ethnic relations, and sociology of education. His research on education focuses on the determinants of underachievement among Hispanic youth in schools and the labor market. He has been honored by fellowships from the National Science Foundation, Rockefeller Foundation, Social Science Research Council, and National Academy of Education. He has served on numerous advisory committees for various foundations and federal agencies in addition to three committees of the National Research Council. He received a B.A. degree from Harvard College and M.A. and Ph.D. degrees from the University of Chicago in sociology.

DOROTHY M. GILFORD is director of the National Research Council's Board on International Comparative Studies in Education. Formerly she served as director of the National Center for Education Statistics and as director of the mathematical sciences division of the Office of Naval Research. Her interests are in research program administration, organization of statis-

tical systems, and education statistics. A fellow of the American Statistical Association, she has served as vice president of the association, chair of its committee on fellows, and chair of its committee on international relations in statistics. She is a member of the International Statistics Institute. She received B.S. and M.S. degrees in mathematics from the University of Washington.

JAMES W. GUTHRIE is codirector of Policy Analysis for California Education (PACE) and professor of education at the University of California at Berkeley. Before joining the University of California faculty, Guthrie was the education specialist in the United States Senate. His research interests include educational policy, school finance and governance, and the reform of educational systems. His publications include books on school finance, educational administration, strategic planning, and teacher education. He has served as vice president of the American Educational Research Association and currently edits that organization's journal, *Educational Evaluation and Policy Analysis*. He has been honored as an Alfred North Whitehead postdoctoral fellow at Harvard University, visiting fellow at the department of educational studies of Oxford University, and in 1990 was named as the American Education Research Association's first senior fellow. He received B.A., M.A., and Ph.D. degrees from Stanford University.

EDWARD H. HAERTEL is associate professor of education at Stanford University. His research interests center on psychometrics and assessment policy, especially large-scale assessments. He has worked with the National Center for Education Statistics; the National Assessment Governing Board; the Center for Research on Evaluation, Standards, and Student Testing; and other public and private organizations on issues concerning the National Assessment of Educational Progress, and with the state of California on issues concerning the California Assessment Program. He received a B.A. degree in mathematics from the University of Wisconsin and a Ph.D. in education from the University of Chicago.

MORRIS H. HANSEN (deceased, October 1990) was chair of the board of Westat, Inc., and an authority on sample survey

design and the conduct of surveys. He joined Westat in 1968 and took a lead role in the design and conduct of many surveys and on the application of statistical quality control methods in welfare programs. Formerly he was associate director of the Bureau of the Census, where he was responsible for the research and development program. He served as president of the Institute of Mathematical Statistics, the American Statistical Association, and the International Association of Survey Statisticians. He was an honorary fellow of the Royal Statistical Society and a member of the National Academy of Sciences. He received a B.S. degree in accounting from the University of Wyoming, an M.A. degree in statistics from American University, and an honorary degree of doctor of laws from the University of Wyoming.

STEPHEN P. HEYNEMAN is the division chief for population and human resources in the technical department in the Europe, Middle East, and North Africa Region of the World Bank. His division's responsibility includes policies on education, health, family planning and nutrition. He has been using international surveys of academic achievement to study problems of educational quality since his first project in Uganda in 1970, and he has worked on educational problems in many developed and developing countries. His other research interests include higher education finance and management, policy reforms in vocational education, examination reforms and standardized testing, cognitive skills, and economic productivity. He is the president-elect of the Comparative and International Education Society. He received a B.A. degree from the University of California at Berkeley, an M.A. degree from the University of California at Los Angeles, and a Ph.D. degree from the University of Chicago in comparative education.

DANIEL G. HORVITZ is distinguished institute scientist at the Research Triangle Institute where he has also been executive vice president and vice president for statistical sciences. His research interests are in survey design and methods with emphasis on cost-effective ways to improve data quality. He is a fellow of the American Association for the Advancement of Science and the American Statistical Association, which he has served as vice president, and a member of the International

Statistical Institute. He received a B.S. degree from the University of Massachusetts and a Ph.D. degree in statistics from Iowa State University.

LYLE V. JONES is professor of psychology and director of the L. L. Thurstone Psychometric Laboratory at the University of North Carolina, Chapel Hill, where he has served as vice chancellor and dean of the graduate school. Formerly he was a faculty member at the University of Chicago and has held visiting faculty appointments at the universities of Illinois, Texas, and Washington. He is a member of the Institute of Medicine, a fellow of the American Academy of Arts and Sciences, and a former president of the Association of Graduate Schools, the Psychometric Society, and the Division of Evaluation and Measurement of the American Psychological Association. He has served as a member of several boards, commissions, and committees of the National Research Council. Among his recent publications are several that focus on differential trends in U.S. school achievement for black and white students. He received B.S. and M.S. degrees from the University of Washington and a Ph.D. degree from Stanford University.

LAURA LATHROP is research assistant to the board. She has served as an educational consultant to various organizations abroad as well as to the U.S. Department of Labor and the University of Maryland. She received a B.A. in psychology from Hollins College, an M.A. degree in education from Inter-American University in Puerto Rico, and a Ph.D. degree in human development and anthropology from the University of Maryland.

GAEA LEINHARDT is a senior scientist at the Learning Research and Development Center and professor of education at the University of Pittsburgh. She has directed projects on program evaluation, teacher assessment, and teacher expertise in mathematics and history. Her research interests focus on classroom instruction and learning in specific subject areas as well as in understanding how instructional explanations are constructed and when they are effective. She has worked on the design of innovative assessment techniques for teachers and students for the National Board of Professional Teacher Standards and sev-

eral state departments of education. She received a B.A. degree in Russian civilization and an M.S. degree in urban education from the University of Chicago and a Ph.D. degree in research methodology from the University of Pittsburgh.

JOHN SCHWILLE is professor and assistant dean for international studies at the College of Education at Michigan State University. His experience includes qualitative as well as quantitative educational research in both international and U.S. domestic settings. His involvement in international assessment dates from 1972 when he received a Spencer Fellowship at Stockholm University for analysis of data collected by the International Association for the Evaluation of Educational Achievement. His most recent field research has dealt with primary school effectiveness issues in Burundi. He received a B.A. degree in the history of France from Harvard University and a Ph.D. degree in comparative sociology of education from the University of Chicago.

FLORALINE I. STEVENS is director of the Program Evaluation and Assessment Branch for the Los Angeles Unified School District and adjunct assistant professor of tests and measurements at the California State University, Los Angeles. Her interests are program evaluation and research issues focused on at-risk students. She served as vice president for the division on School Evaluation and Program Development of the American Educational Research Association (AERA), chair of the Research Steering Committee for the Council of Great City Schools, and chair of the School Research Directors at AERA. She currently serves on the National Indicators Panel for the U.S. Department of Education. She received a B.S. degree from the University of Southern California and M.Ed. and Ed.D. degrees from the University of California at Los Angeles.

JUDITH TORNEY-PURTA is professor of human development and affiliate professor of psychology at the University of Maryland, specializing in applied developmental and educational psychology. Formerly she was professor of psychology at the University of Illinois at Chicago. She was the senior author of the International Association for the Evaluation of Educational Achievement volume reporting the 10-country study of civic education.

She has served on the Fulbright Senior Scholar Selection Panel in Psychology and as chair of the International Relations Committee of the American Educational Research Association, and she is a fellow of the American Psychological Association. Her current research interests are in applying cognitive modeling techniques to measure the complexity and constraints in adolescents' views of the political and economic world. She received an A.B. degree in psychology from Stanford University and M.A. and Ph.D. degrees in human development from the University of Chicago.

DAVID E. WILEY is professor and dean at the School of Education and Social Policy, Northwestern University. A statistician and psychometrician by training and early work, much of his recent research and writing has focused on public policy as related to educational testing, teaching-learning processes, and legislative initiatives affecting these aspects of education. He has been involved in international comparative studies of education since 1971 and recently completed (with T. N. Postlethwaite) a volume reporting findings of the second science study of the International Association for the Evaluation of Educational Achievement (IEA). He also serves on the IEA International Technical Committee. His current research is focused on integrating frameworks for the assessment of learning, ability, and performance. He received a A.B. degree from San Diego State College, and M.S. and Ph.D. degrees from the University of Wisconsin.

RICHARD M. WOLF is professor of psychology and education and chair of the Department of Measurement, Evaluation, and Statistics at Teachers College, Columbia University. His current work is in the areas of international research and educational evaluation. He is currently an individual member of the International Association for the Evaluation of Educational Achievement and serves as chair of its editorial committee and member of its technical advisory committee and steering committee for the computers in education project. He received a B.A. degree from Antioch College, an Ed. M. degree from the University of Buffalo, and a Ph.D. degree from the University of Chicago.

NATIONAL ACADEMY PRESS

The National Academy Press was created by the National Academy of Sciences to publish the reports issued by the Academy and the National Academy of Engineering, the Institute of Medicine, and the National Research Council, all operating under the charter granted to the National Academy of Sciences by the Congress of the United States.